AD_____

Award Number: W81XWH-12-1-0561

TITLE: Development of the Ovarian Cancer Cohort Consortium: Risk Factor Associations by Heterogeneity of Disease

PRINCIPAL INVESTIGATOR: Dr. Shelley Tworoger

CONTRACTING ORGANIZATION: Brigham and Women's Hospital, Inc.

Boston, MA 02115-6110

REPORT DATE: October-2013

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October-2013	Annual	30September2012-29September2013

4. TITLE AND SUBTITLE
Development of the Ovarian Cancer Cohort Consortium: Risk Factor Associations by Heterogeneity of Disease

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-12-1-0561

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dr. Shelley Tworoger

E-Mail: nhsst@channing.harvard.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Brigham and Women's Hospital, Inc.
75 Francis Street
Boston, MA 02115-6110

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The objective of this application is to study the etiologic heterogeneity of ovarian cancer in multiple cohorts and to build the infrastructure of the Ovarian Cancer Cohort Consortium (OC3), an international consortium of cohort studies, to address scientific aims important for understanding ovarian cancer risk, early detection, and tumor heterogeneity that are only feasible in a consortium setting. Specifically we will examine associations of risk factors with invasive ovarian cancer, including (but not limited to) age, OCs, tubal ligation, parity, postmenopausal hormone use, family history of ovarian cancer, BMI, height, analgesic use, and lifetime ovulatory cycles, differ by histologic subtype, tumor dominance (as a surrogate for cell of origin), and tumor aggressiveness (tumors fatal within three years vs. all others). Then we will determine if risk prediction models for ovarian cancer can be improved by accounting for differential associations by cancer phenotype. In addition, the proposed efforts will create an infrastructure with a core dataset of important variables for ovarian cancer epidemiology that will be available for future efforts to study ovarian cancer risk, including projects that will use prospectively collected biological specimens. Currently, 22 cohorts have agreed to participate in the OC3. We have executed data use agreements between the Brigham and Women's Hospital (data coordinating center) with 15 cohorts; 4 cohorts do not require an agreement. We have received data from 17 cohorts, with 2 cohorts actively preparing data. Once the data use agreements have been finalized for the final 3 cohorts, we will work to receive those data. Data harmonization is nearly complete for the cohorts for which we have received data. Preliminary analyses of primary ovarian cancer risk factors (e.g., oral contraceptive use, parity) are on-going to ensure that the expected associations are observed and to determine whether additional data cleaning is needed. The goal is to receive data from the remaining cohorts by the end 2013 and complete the analysis with the OCAC and the risk factor associations by histology by early to mid 2014. Several projects have been proposed to utilize this resource and attempts to obtain funding are underway.

15. SUBJECT TERMS
Ovarian Cancer, tumor heterogeneity, histology, cell of origin, tumor aggressiveness, risk prediction

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	37	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Progress Report for OC110197
Sept. 2013

INTRODUCTION

The objective of this translational leverage award is to study the etiologic heterogeneity of ovarian cancer in multiple cohorts and to build the infrastructure of the Ovarian Cancer Cohort Consortium (OC3). The OC3 is an international consortium of cohort studies designed to address scientific aims important for understanding ovarian cancer risk, early detection, and tumor heterogeneity. The OC3 is part of the NCI Cohort Consortium, which is an extramural-intramural partnership to address the need for large-scale collaborations and provides the super-structure (but not funding) for managing the OC3. The OC3 currently has 22 participating, on-going cohort studies and we expect there to be over 8,000 ovarian cancer cases among more than 1.5 million women. The goals of the OC3 are to bring together cohorts with ovarian cancer endpoints for pooled projects, build a focused group of ovarian cancer researchers, and develop a comprehensive approach that integrates questionnaire and pathology data with biomarkers, genetics, and tissue. In addition to building the OC3 infrastructure, we propose to evaluate associations of ovarian cancer risk factors by different metrics of tumor heterogeneity. The first specific aim of this application is to examine whether associations of known and putative ovarian cancer risk factors, including (but not limited to) age, oral contraceptive use, tubal ligation, parity, postmenopausal hormone use, family history of ovarian cancer, body mass index, height, analgesic use, and lifetime ovulatory cycles, differ by (a) histologic subtype, (b) tumor dominance (as a surrogate for cell of origin), and (c) tumor aggressiveness (tumors fatal within three years vs. all others). We will use the information generated in the first aim to conduct the second specific aim, which is to develop ovarian cancer risk prediction models accounting for differential associations by cancer phenotype.

BODY

This grant began on September 30, 2012. Currently, 22 cohorts have agreed to participate in projects addressing the risk factor associations by tumor heterogeneity and to develop an improved risk prediction model for ovarian cancer. The tasks completed in the first year included: (1) identifying interested cohorts and formalizing the data dictionary, (2) circulating the list of desired variables to participating cohorts, (3) having cohorts prepare data and send to the Brigham and Women's Hospital (BWH) data coordinating center (DCC) for harmonization, (4) when available, conducting additional pathologic abstraction, and (5) beginning to harmonize variables centrally at the DCC.

A data dictionary was developed by the OC3 Steering Committee, which includes the PI of this grant as well as subcontract PIs, Drs. Kala Visvanathan and Alan Arslan. The data dictionary and a short questionnaire about the data collection and attributes were sent to all interested cohorts (see Appendix 1). We also developed a flow chart for abstracting tumor dominance from pathology reports (Appendix 2). Only a subset of 10 cohorts have collected pathology reports – we currently are working with these cohorts to have personnel from the DCC either train on-site staff for abstraction or to travel to the site and conduct the abstraction directly. Due to the difficulties in obtaining the appropriate permissions to access the pathology reports, this work will continue into year 2.

In total, 22 cohorts from the US, Australia, Europe, and Asia have agreed to participate. For IRB purposes we require that each cohort establish a data use agreement (DUA) or provide a letter stating that the IRB does not require a DUA (if sending completely de-identified data). Three cohorts are primed at the BWH and therefore are covered under the primary IRB protocol and thus do not need a DUAA. We have received executed DUAs between the BWH and 15 cohorts; 1 cohort did not require an agreement. Negotiations are on-going for the remaining 3 cohorts. We have received data from 17 cohorts, with 2 cohorts actively preparing data. The goal is to receive data from the remaining cohorts by the end 2013. Details of the participating cohorts and their status in the OC3 as well as estimated or actual sample sizes are presented in Table 1.

Data harmonization for the key variables to be used in this analysis is complete for 13 cohorts from which we have received data; cleaning is ongoing for 4 studies. Specifically we have focused on cleaning and harmonizing the following variables: ovarian cancer diagnosis characteristics (date/age of diagnosis, date of death, type of tumor, morphology, histology, grade), study enrolment and follow-up data (date/age of enrolment,

date/age of death, date/age of last follow-up), race, prior cancer diagnoses, family history of ovarian or breast cancers, menopausal status, postmenopausal hormone use (ever/never, duration, and type), use of oral contraceptives (ever/never, duration), tubal ligation, parity, hysterectomy status, oophorectomy status, age at menarche, age at menopause, smoking, height, body mass index (BMI), BMI at age 18, alcohol intake, and NSAID/aspirin use. Other variables collected through the OC3 will be cleaned at a later time when additional funding is obtained. However, we felt it best to obtain a wide variety of potential variables of interest to facilitate future studies.

In evaluating existing risk prediction models that could be used as the comparison model for the risk prediction aim, it became clear that existing studies have been based on small sample sizes or had limited validation. Therefore we decided that it was crucial to develop a "base" risk prediction model that could be used as a comparison model to evaluate improvements from multiple perspectives (e.g., incorporating tumor heterogeneity or genetics) with a large sample size and a statistically rigorous approach. To establish such a model, we are collaborating with the Ovarian Cancer Association Consortium (OCAC), which is a consortium of case-control studies, to develop and validate a risk prediction model. The over-arching plan for this is presented in Figure 1. In summary, the OCAC data will be used to develop the initial input relative risks for key ovarian cancer risk factors as well as to develop imputation models to deal with missing data using US women. The risk estimates will be stratified by age at diagnosis (<50, ≥50) and parity (nulliparous, parous) as these are potential modifying factors. In addition, other inputs to the final model will include prevalence of bilateral salpingo-oophorectomy (BSO) in the US population by birth cohort and age using NHANES data, the relative risk of ovarian cancer after BSO using a meta-analysis of existing and unpublished literature (based on 8 studies, RR=0.11, 95%CI=0.07-0.18), and ovarian cancer incidence rates in the US population using SEER (which are not corrected for BSO rates). These values will be used to determine the baseline absolute risk in the population. Then, the risk prediction model developed in the OCAC studies will be evaluated in the US-based OC3 studies, holding out two cohorts for independent validation; the model will be refined and finally validated in the two independent cohorts. The initial phase of the work in the OCAC is nearly complete and then analyses in the OC3 will begin this fall, with the final project to be completed in early 2014. This unique collaboration will provide a resource for all future work on ovarian cancer risk prediction.

We have developed SAS macros for conducting analyses in a standardized manner, including a macro to meta-analyze results for a particular exposure across studies, one to conduct a pooled analysis, and macros to assess risk factor association heterogeneity by tumor subtype. Preliminary analyses of primary ovarian cancer risk factors (e.g., oral contraceptive use, parity) with total ovarian cancer are on-going to ensure that the expected associations are observed and to assess potential heterogeneity across studies. We see the expected associations for oral contraceptive use and parity within the 13 cohorts with clean data (Table 2). For example, ever use of oral contraceptives was associated with a 12% decrease in ovarian cancer risk (95%CI=0.80-0.96); however there was evidence of heterogeneity between studies (Q statistic=23.9, p=0.03). Similar results were observed for OC duration (RR, per 1 year of use=0.98, 95%CI=0.97-0.99, Q statistic=47.7 with a p<0.001). We currently are exploring the potential reasons for this heterogeneity (e.g., different birth cohorts, age at baseline or diagnosis) and whether this heterogeneity exists when examining oral contraceptive duration among ever users. We hope to identify the source of variability because we will need to conduct a pooled analysis for the rarer histologic subtypes. For parity, we observed 29% reduction in risk for parous versus nulliparous women (95%CI=0.62-0.80, Q statistic=17.5 with a p=0.09) and a 10% reduction in risk per pregnancy (95%CI=0.88-0.93, Q statistic=19.6 with a p=0.11). We also have evaluated the sample size for different tumor subtypes, based on the data from the 13 studies we currently have cleaned in our database (n=4,177 cases). At this time we have 2,270 serous, 406 endometrioid, 233 mucinous, 178 clear cell, 145 poorly differentiated, and 945 other/unknown subtypes. We expect to have a total of 8,000 cases, so these numbers likely will increase by at least 50%. Our goal is to complete the analysis of risk factor associations by histology by early to mid 2014.

With respect to the OC3 structure, we continue to have monthly conference calls run by the PI with the Steering Committee (Table 3). At this time, the calls focus on obtaining data and other details regarding study participation, discussing potential future collaborations or projects, and vetting preliminary results. The Steering Committee has developed publication guidelines (Appendix 3) and provided feedback on a template

DUA for participating cohorts and those who plan to use the data for future projects (Appendix 4). The PI also meets weekly with Dr. Elizabeth Poole (a junior faculty member working on the project) and the OC3 programmer. The OC3 has had two in-person meetings since the grant started, one in October 2012 at the NCI Cohort Consortium annual meeting and one in April 2013 at the AACR Annual Meeting. Our next in-person meeting is in November 2013 at the upcoming Cohort Consortium annual meeting. We chose these meeting times because many investigators attend these associated meetings so we have very good attendance. We also are developing a website for the OC3 to communicate our goals, guidelines for participation, and in the future, interesting findings from the study (see https://sites.google.com/a/channing.harvard.edu/oc3/?pli=1).

Seven projects have been proposed using the infrastructure of the OC3 in addition to the aims funded by this grant (Table 4). These projects will obtain separate funding to support the analyses needed for that project or any additional data collection. For example, Dr. Elizabeth Poole (key personnel on this award) submitted an R03 to the National Cancer Institute (NCI) to examine inflammatory exposures and risk of ovarian cancer. The grant received an impact score of 32 and will be resubmitted in November 2013. Other investigators are applying for intramural NCI funding or funding from the European Union. We also plan to participate in the NCI-initiated OncoArray project, which will conduct GWAS plus assay other SNPs of interest on multiple types of cancer, including ovarian cancer. Nine OC3 cohorts with appropriate biospecimens plan to participate and send samples using a nested case-control study design. The OC3 will coordinate data management with the OCAC (who is managing the overall ovarian OncoArray project) and will add the genotyping data to the OC3 database after the initial analyses are complete within the full OncoArray consortium (genotyping and data cleaning expected to be complete at the end of 2014).

KEY RESEARCH ACCOMPLISHMENTS

Below is a list of key research accomplishments in the first year of this award.

- Recruited 22 cohorts to the OC3 including establishing 15 DUAs (4 cohorts do not need a DUA); DUA negotiations are on-going for 3 cohorts
- Developed a data dictionary to provide guidelines for sending data
- Received data from 17 cohorts to date
- Developed coding guidelines for tumor dominance
- Developed collaborative structure and framework for the OC3
- Completed data harmonization for 13 cohorts, with 4 cohorts in the pipeline
- Completed and tested SAS macros for data analysis
- Developed of a project plan for the risk prediction model project with the OCAC and for the initial analysis by histology
- Continued participation by OC3 cohorts and investigators in monthly steering committee meetings and bi-annual in person meetings
- Multiple projects proposed to use OC3 resource with additional funding being sought

REPORTABLE OUTCOMES

At this time, the primary reportable outcome is the development of the OC3 database, which contains data on ovarian cancer risk factors and outcomes from 17 cohort studies and by the end of 2013 will contain data from 5 more studies. This resource can be used for the analyses proposed in this grant as well as other analyses (see Table 4 for projects that have been proposed to date). In addition, several investigators have applied for funding to the NIH/NCI (Poole), NCI intramural funds (Wentzensen), and European Union funds (Lukanova) using the OC3 infrastructure.

CONCLUSION

We are actively developing the OC3 infrastructure by pooling existing cohort data to better elucidate the biology of ovarian cancer. Scientifically, we will evaluate whether associations for putative ovarian cancer risk factors differ by tumor subtypes (histology, cell of origin, aggressiveness), as well as develop risk prediction models based on differing risks across subtypes. Further, we are working to develop a “base” risk prediction model that can be used as a comparison for assessing improvement in future work. This will be beneficial to the entire ovarian cancer research community. This systematic approach to address ovarian cancer heterogeneity in a large consortial effort will set new standards for evaluating ovarian cancer risk factors and

biomarkers and thereby impact understanding of ovarian cancer etiology beyond the work conducted in OC3. Further, the classification of ovarian cancers by histology, cell of origin, and aggressiveness in a large set of cohort studies will set an important harmonized framework for future risk and biomarker studies of ovarian cancer. This project will make extremely efficient use of existing data to produce new information that would otherwise not be available because of limited statistical power within individual studies as well as provide a basis for future consortial studies in the OC3.

REFERENCES

No manuscripts from the OC3 have been published at this time.

APPENDICES

Appendix 1: OC3 data dictionary and data collection questionnaire
Appendix 2: Tumor dominance coding flowchart
Appendix 3: Proposal and publication guidelines for the OC3
Appendix 4: Template data use agreements for participating cohorts and for those wishing to access the OC3 database with approved projects

SUPPORTING DATA

Table 1: Participating cohorts in the OC3 and estimated numbers of women and ovarian cancer cases as well as the status of the data use agreement (DUA) and receipt of data to the data coordinating center (DCC)
Figure 1: Schematic for developing and validating a risk prediction model of ovarian cancer in the OCAC and OC3
Table 2: Preliminary results for the association of oral contraceptive use and parity with total ovarian cancer risk among 13 of the cohorts participating in the OC3
Table 3: OC3 Steering Committee members
Table 4: Additional proposed projects using the OC3 infrastructure and the status

APPENDIX 1: OC3 data dictionary and data collection questionnaire			
Variable Name	**Description**	**Coding**	**Comment**
ID	Unique ID for each study participant	Please do not use the real participant ID, but assign an alias or fake ID	If you need assistance assigning fake IDs, please let us know.
CASE CHARACTERISTICS			
OVCA	Ovarian cancer case identifier	1=epithelial ovarian cancer case; 2=peritoneal case; 3=fallopian tube case; 4=germ cell case; 5=sex cord case; missing if non-case	
BEHAVIOR	Borderline or invasive case	1=borderline; 2=invasive; 9=missing; missing if non-case	
HOWCONF	How ovarian cancer case was confirmed	1=self-report only; 2=pathology report review; 3=pathology slide review; 4=registry; 5=other; 9=missing; missing if non-case	
HISTOLOGY	Identifier of case histology	1=serous; 2=endometrioid; 3=mucinous; 4=clear cell; 5=poorly differentiated; 6=other; 9=unknown; missing if non-case	Fill in if case confirmed through pathology review; otherwise, see ICD histology variable below.
ICD9CODE	ICD9 code for cases	As coded, missing if non-case	If available
ICDOCODE	ICD-O code for cases	As coded, missing if non-case	If available
ICDOHIST	ICD-O histology code for cases	As coded, missing if non-case	If available
STRUCTURE	Indicator of which ovarian structure was measured in pathology report	1=tumor; 2=ovary; 3=adnexa; 998=case was clearly dominant or non-dominant, no measurements required; 999=no measurements listed; missing if non-case	
DOMINANCE	Identifier of tumor dominance	1=dominant on right; 2=dominant on left; 3=non-dominant; 9=unknown; missing if non-case	
RDIM1	1st measurement of the tumor/ovary/adnexa on right side, if given	in cm; 998=case was clearly dominant or non-dominant, no measurements required; 999=no measurements listed; missing if non-case	if 3 dimensions not given, fill in as many as available in path report
RDIM2	2nd measurement of the tumor/ovary/adnexa on right side, if given	in cm; 998=case was clearly dominant or non-dominant, no measurements required; 999=no measurements listed; missing if non-case	
RDIM3	3rd measurement of the tumor/ovary/adnexa on right side, if given	in cm; 998=case was clearly dominant or non-dominant, no measurements required; 999=no measurements listed; missing if non-case	

LDIM1	1st measurement of the tumor/ovary/adnexa on left side, if given	in cm; 998=case was clearly dominant or non-dominant, no measurements required; 999=no measurements listed; missing if non-case	if 3 dimensions not given, fill in as many as available in path report
LDIM2	2nd measurement of the tumor/ovary/adnexa on left side, if given	in cm; 998=case was clearly dominant or non-dominant, no measurements required; 999=no measurements listed; missing if non-case	
LDIM3	3rd measurement of the tumor/ovary/adnexa on left side, if given	in cm; 998=case was clearly dominant or non-dominant, no measurements required; 999=no measurements listed; missing if non-case	
DXYEAR	Year of ovarian cancer diagnosis	Year of diagnosis as YYYY; missing if non-case	See instructions tab: only provide month and year if your institution allows you to send a "limited" dataset.
DXMONTH	Month of ovarian cancer diagnosis	Month of diagnosis as MM; missing if non-case	If available
AGEDX	Age at ovarian cancer diagnosis	Age at diagnosis in months; missing if non-case	Provide age if year and month of diagnosis are not available
STAGE	Tumor stage	0=in situ; 1=localized; 2=regional; 3=distant; 9=unstaged or DK; missing if non-case	
GRADE	Grade	1=well differentiated; 2=moderately differentiated; 3=poorly differentiated; 4=undifferentiated; 9=DK; missing if non-case	If you do not feel that the grade classification you received is accurate, please do not provide it.
SOURCE	Source of information about tumor characteristics	1=SEER; 2=pathology report review; 3=pathology slide review; 4=other; missing if non-case	
INFORMATION ABOUT OTHER DISEASES IN THE COHORTS			
SYNCCA	Was the ovarian cancer diagnosed within 3 months as another cancer?	0=no; 1=yes; 9=unknown; missing if not an ovarian cancer case	For other cancers diagnosed during follow-up, please included self-reported as well as pathology confirmed cancers.
SYNCTYPE	What type of cancer was the synchronous cancer	1=breast; 2=colorectal; 3=endometrial; 4=other; 9=unknown; missing if 0 or . for SYNCCA	

BASECA	Diagnosed with any other non-melanoma skin cancer before baseline	0=no; 1=yes; 9=unknown	
OTHCA	Diagnosed with any other non-melanoma skin cancer during follow-up	0=no; 1=yes; 9=unknown	
OCADXYEAR	Year of other cancer diagnosis	Year of diagnosis as YYYY; missing if non-case; 9999 if unknown	See instructions tab: only provide month and year if your institution allows you to send a "limited" dataset.
OCAAGE	Age of other cancer diagnosis	Age at diagnosis as YY; missing if non-case; 999 if unknown	To be provided if year of diagnosis is unavailable
CVD	Ever diagnosed with cardiovascular disease (do not include angina) at baseline	0=no; 1=non-fatal MI; 2=fatal MI; 3=stroke; 4=bypass surgery; 5=angioplasty	
CVDYEAR	Year diagnosed with CVD	Year of diagnosis as YYYY; missing if non-case	See instructions tab: only provide month and year if your institution allows you to send a "limited" dataset.
CVDAGE	Age at diagnosis of CVD	Age at diagnosis as YY; missing if non-case; 999 if unknown	To be provided if year of diagnosis is unavailable
DIABETES	Ever diagnosed with diabetes at baseline	0=no; 1=yes; 9=unknown	
DBYEAR	Year diagnosed with diabetes	Year of diagnosis as YYYY; missing if non-case; 9999 if unknown	See instructions tab: only provide month and year if your institution allows you to send a "limited" dataset.
DBDXAGE	Age at diagnosis of diabetes	Age at diagnosis as YY; missing if non-case; 999 if unknown	To be provided if year of diagnosis is unavailable
DBINS	Taking insulin for diabetes management at baseline	0=no; 1=yes; 9=unknown; missing if not a diabetes case	
AID	Ever diagnosed with auto-immune disease at baseline	0=no; 1=lupus; 2=rheumatoid arthritis; 3=Graves' Disease; 4=inflammatory bowel disease; 5=scleroderma; 6=multiple sclerosis; 7=other	
AIDYEAR	Year diagnosed with auto-immune disease	Year of diagnosis as YYYY; missing if non-case	
AIDAGE	Age at auto-immune disease diagnosis	Age at diagnosis as YY; missing if non-case; 999 if unknown	To be provided if year of diagnosis is unavailable

PID	Ever diagnosed with pelvic inflammatory disease at baseline	0=no; 1=yes; 9=unknown	
PIDYEAR	Year diagnosed with pelvic inflammatory disease	Year of diagnosis as YYYY; missing if non-case	
PIDAGE	Age at pelvic inflammatory disease diagnosis	Age at diagnosis as YY; missing if non-case; 999 if unknown	To be provided if year of diagnosis is unavailable
ENROLLMENT AND FOLLOW-UP INFORMATION			
QXYEAR	Year of baseline questionnaire return	Year of baseline questionnaire return as YYYY	See instructions tab: only provide month and year if your institution allows you to send a "limited" dataset.
QXMONTH	Month of baseline questionnaire return	Month of baseline questionnaire return as MM	
QXAGE	Age at baseline questionnaire return	Age at baseline questionnaire return as YY; 999 if unknown	
DEATH	Indicator of whether alive or dead at end of follow-up	0=alive; 1=dead; 9=unknown	
DEATHYEAR	Year of death	Year of death as YYYY; missing if not dead by end of follow-up; 9999 if unknown	See instructions tab: only provide month and year if your institution allows you to send a "limited" dataset.
DEATHMONTH	Month of death	Month of death as MM; missing if not dead by end of follow-up; 99 if unknown	
DEATHAGE	Age at death	Age at death as YY; missing if not dead by end of follow-up; 999 if unknown	
LASTYEAR	Year of last follow-up	Year of last follow-up as YYYY	See instructions tab: only provide month and year if your institution allows you to send a "limited" dataset.
LASTMONTH	Month of last follow-up	Month of last follow-up as MM	
LASTAGE	Age at last follow-up	Age at last follow-up as YY; 999 if unknown	
DEMOGRAPHIC DATA			

BIRTHYEAR	Year of birth	Year of birth as YYYY	If year of birth is not available, please provide age at enrollment.
BIRTHMONTH	Month of birth	Month of birth as MM	If IRB does not allow this to be sent, please include age at baseline and year of birth only
RACE	Race	1=White; 2=Black; 3=Asian/Pacific Islander; 4=Native American; 5=Other; 9=Unknown	
HISPANIC	Ethnicity	1=Hispanic; 0=Non-Hispanic	
EDUCTION	Highest level of education	1=did not finish high school; 2=high school; 3=some college; 4=completed college; 5=post graduate; 9=unknown/missing	
ANTHROPOMETRIC VARIABLES			
HEIGHT	Height in inches	999=unknown	
WEIGHT18	Weight at age 18 in pounds	Weight; 999=unknown	
WEIGHT	Weight in pounds	999 if unknown	
WAIST	Waist circumference in inches	999=unknown	
HIP	Hip circumference in inches	999=unknown	
MENARCHE AND MENOPAUSE VARIABLES			
AGEMENARCHE	Age when menstrual periods began	99=unknown	
IRREG	Ever had irregular periods	0=no, usually regular; 1=sometimes irregular; 2=frequently irregular	
MENLENGTH	Average length of menstrual cycle	Length of menstrual cycle in days	
MENO	Menopause status	1=post; 0=pre; 9=unknown	
RMENO	Reason for menopause	0=pre; 1=natural; 2=radiation; 3=surgery; 4=other; 9=unknown	
AGEMENO	Age at natural menopause	Missing if pre or had uterus removed before menopause; 999=if menopausal; but unknown age	
REPRODUCTIVE HISTORY			
PREG	Number of pregnancies lasting > 6 months	Twins count as a single birth; 999=unknown	
AGEFIRST	Age at first birth	Age in months; 999=unknown/missing	

B		
AGELASTB	Age at last birth	Age in months; 999=unknown/missing
DURBF	Duration of breastfeeding for all children combined	Duration in months; 999=unknown; missing if never parous
DURBF1	Duration of breastfeeding for 1st child	Duration in months; 999=unknown; missing if never parous
DURBF2	Duration of breastfeeding for 2nd child	Duration in months; 999=unknown; missing if never parous
DURBF3	Duration of breastfeeding for 3rd child	Duration in months; 999=unknown; missing if never parous
DURBF4	Duration of breastfeeding for 4th child	Duration in months; 999=unknown; missing if never parous
DURBF5	Duration of breastfeeding for 5th child	Duration in months; 999=unknown; missing if never parous
INFERT	Infertility ; ever tried to get pregnant for one year without success?	0=No; 1=Yes - due to female causes; 2=Yes - due to male causes; 3=Yes; cause unknown; 9=unknown
INFERTTX	Ever treated for infertility	0=No; 1=ovulation induction; 2=IVF; 9=unknown
CONTRACEPTION HISTORY		
TUBAL	Tubal ligation	1=reported having had a tubal ligation; 0=no report of tubal ligation; 9=unknown/missing
TUBALTYPE	Type of tubal ligation	1=ligation; 2=ring/band; 3=cauterization; 9=unknown/missing; missing if no tubal
TUBALYEAR	Year of tubal ligation	Year of tubal ligation as YYYY; 9999=missing
TUBALAGE	Age at tubal ligation	Age at tubal ligation in months; 999=missing
OCUSE	Ever use of oral contraceptives	1=reported ever using OCs; 0=reported never using OCs; 9=unknown/missing
OCDUR	Duration of OC use	Duration in years; 0=non-user; 999=unknown duration

OCAGEFIRST	Age when began using OCs	Age in years; missing if never user; 999=unknown age among users
OCAGELAST	Age when stopped usign OCs	Age in years; missing if never user; 999=unknown age among users
IUD	Ever use of IUDs	1=reported ever using IUDs ; 0=reported never using IUDs; 9=unknown/missing
PATCHCONT	Ever use of patch contraceptives	1=reported ever using patch contraceptives; 0=reported never using patch contraceptives; 9=unknown/missing
FOAM	Ever use of foam or jelly contraceptives	1=reported ever using foam or jelly contraceptives; 0=reported never using foam or jelly contraceptives; 9=unknown/missing
MALECONT	Ever use of male contraceptives (i.e. condom)	1=reported ever using male contraceptives; 0=reported never using male contraceptives; 9=unknown/missing
OTHERCONT	Ever use of other contraceptives	1=reported ever using other contraceptives; 0=reported never using other contraceptives; 9=unknown;missing

FAMILY HISTORY OF BREAST AND/OR OVARIAN CANCER		
HXBRCA	1st degree family history of breast cancer	1=mother; sister; or daughter had breast cancer ; 0=no 1st degree family history; 9=unknown
BRCANUM	Number of 1st degree relatives with breast cancer	Missing if no family history
HXOVCA	1st degree family history of ovarian cancer	1=mother; sister; or daughter had ovarian cancer ; 0=no 1st degree family history; 9=unknown
OVCANUM	Number of 1st degree relatives with ovarian cancer	Missing if no family history

HYSTERECTOMY/OOPHORECTOMY STATUS		
HYST	Hysterectomy status	0=No; 1=Yes; 9=Unknown
HYSTYEAR	Year of hysterectomy	Year of hysterectomy as YYYY; 9999=missing/unknown
HYSTAGE	Age at hysterectomy	Age at hysterectomy in months; 999=missing/unknown
OOPH	Oophorectomy status	0=No; 1=Yes - one ovary removed; 2=Yes - both ovaries removed; 3=Yes - unknown ovaries removed; 9=unknown

OOPHYEAR	Year of oophorectomy	Year of oophorectomy as YYYY; 9999=missing/unknown
OOPHMONTH	Month of oophorectomy	Month of oophorectomy as MM; 99=missing/unknown
OOPHAGE	Age at oophorectomy	Age at oophorectomy in months; 999=missing/unknown
POST-MENOPAUSAL HORMONE USE		
PMH	Use of any type of post-menopausal hormones	0=No; 1=Yes; 9=Unknown; missing if pre
DURPMH	Duration of use of any type of PMH	Duration in years; 999=unknown; missing if pre
EONLY	Ever use of oral estrogen only	0=No; 1=Yes; 9=Unknown; missing if pre
DUREST	Duration of use of oral estrogen only	Duration in years; 999=unknown; missing if pre
EPLUSP	Ever use of oral estogen plus progesterone	0=No; 1=Yes; 9=Unknown; missing if pre
DUREP	Duration of use of oral estrogen plus progesterone	Duration in years; 999=unknown; missing if pre
OTHPMH	Ever use of other post-menopausal hormones (including non-oral)	0=No; 1=Yes; 9=Unknown; missing if pre
DUROTH	Duration of use of other post-menopausal hormones	Duration in years; 999=unknown; missing if pre
OTHER KNOWN AND POTENTIAL OVARIAN CANCER RISK FACTORS		
TALC	Regular use of talc	0=No; 1=Yes - genital; 2=Yes - other; 3=Yes, unknown; 9=unknown/missing
TALCDUR	Duration of talc use	Duration in years; .=never user; 999=missing among users
TALCFREQ	Frequency of talc use	0=No use; 1=<1 per week; 2=1 or more times per week
ENDO	Endometriosis ; ever diagnosed with endometriosis?	0=No self-report; 1=Yes by self-report, but not confirmed; 2=Yes by self-report and confirmed laparoscopically; 9=Unknown
ENDOYEA	Year diagnosed with	Year of endometriosis as YYYY; 9999=missing/unknown

R	endometriosis	
SMOKE	Smoking status	0=Never; 1=Former; 2=Current
PACKYR	Pack years of smoking	999=unknown; missing if never smoker
QUITSMOKE	Year quit smoking	999=unknown; missing if never smoker or current smoker
ASP	Regular use of aspirin	0=No; 1=Yes; 9=unknown/missing
ASPDUR	Duration of aspirin use	Duration in years; .=never user; 999=missing among users
ASPDOSE	Dose of aspirin used per day	Dose in mg; .=never user; 9999=missing among users
ASPFREQ	Frequency of aspirin use	Frequency in times per month; .=never user; 99=missing among users
NSAID	Regular use of non-aspirin NSAIDs	0=No; 1=Yes; 9=unknown/missing
NSAIDDUR	Duration of NSAID use	Duration in years; .=never user; 999=missing among users
NSAIDDOSE	Dose of NSAID used per day	Dose in mg; .=never user; 9999=missing among users
NSAIDFREQ	Frequency of NSAID use	Frequency in times per month; .=never user; 99=missing among users
TYL	Regular use of tylenol/acetominophen	0=No; 1=Yes; 9=unknown/missing
TYLDUR	Duration of tylenol/acetominophen use	Duration in years; .=never user; 999=missing among users
TYLDOSE	Dose of tylenol/acetominophen used per day	Dose in mg; .=never user; 9999=missing among users
TYLFREQ	Frequency of tylenol/acetominophen use	Frequency in times per month; .=never user; 99=missing among users
STATIN	Ever used statins	0=No; 1=Yes; 9=unknown/missing
METFORMIN	Ever used metformin	0=No; 1=Yes; 9=unknown/missing
ACT	Total physical activity in MET-hours/week	999=unknown/missing
WALK	Walking in MET-hours/week	999=unknown/missing
SIT	Hours per week spent	999=unknown/missing

	sitting	
CAFF	Caffeine intake in g/day	9999=unknown/missing
COFFEE	Coffee intake in cups/day	9999=unknown/missing
TEA	Tea intake in cups/day	9999=unknown/missing
ALC	Alcohol intake in g/day	9999=unknown/missing

VARIABLES INDICATING TYPES OF BIOLOGICAL SPECIMENS COLLECTED		
BLOOD	Collected blood samples	0=No; 1=Yes
URINE	Collected urine samples	0=No; 1=Yes
CHEEK	Collected cheek cells	0=No; 1=Yes
BLOCK	Collected tumor blocks	0=No; 1=Yes

INSTRUCTIONS

1. Please fill out the accompanying data submission form, which includes a basic description of your dataset and which variables you are not providing.
2. If your institution allows you to provide a "limited" dataset, please provide year (and month, where available).
   If your institution only allows a "de-identified" dataset, please provide age.
3. If years of diagnosis/other events are not available, please provide age at diagnosis/other event.
4. If we have requested a continuous variable and you only have categorical data, please provide a categorical variable with the name of the variable with "c" at the end.
   For example, if you only have age at menarche in categories, please call your variable "AGEMENARCHEC"
   On the accompanying data submission form, please tell us how your categorical variable was coded and which values each category comprises.

OC3 Data Submission Form – 6/25/2012

Instructions:

All women in your cohort are potentially eligible to be included in the dataset that you supply to the OC3. The following exclusions should be made, although we can also make these exclusions when cleaning and harmonizing the data.

Exclusions:

1. Men
2. Women with a bilateral oophorectomy prior to baseline
3. Women with unknown oophorectomy status at baseline
4. Women with any non-melanoma skin cancer prior to baseline

If your ovarian cancer cases are confirmed through pathology review, please include all cases (whether epithelial or not; invasive and borderline) that have been confirmed. If your ovarian cancer cases are attained through registry linkages, please see the accompanying form that includes the ICD-9 or ICD-O site codes that we would consider to be ovarian cancer. For histology codes, please send all that were coded for the appropriate ICD codes.
At this time, we are only collecting baseline data. However, we plan to do some methods development to include data updates in the cohorts with available data. Therefore, please fill out the survey below and include this form with your data submission.

Part 1. Please fill out the following form with a brief description of the dataset you are submitting to the OC3

1. Cohort/Study name_____
2. PI Name and email address_____
3. Contact person and email address_____
4. How many records are being submitted? _____
5. How many epithelial ovarian cancer cases are included? _____
6. Please provide a brief, overall description of the cohort members that are included in your submission (i.e., age range, date of enrollment, total amount of follow-up time).
7. Please list which variables you are not providing (i.e. variables that were not collected in your study).
8. Are you providing data coded as in the data dictionary or "raw" data coded in the original format? _____.
   If the latter, please provide a separate data dictionary of the variables to allow our programmers to harmonize the data.
9. Please provide a brief description of how the following variables were coded.

ICD codes (if provided, please list classification used)
Stage (i.e. from pathology reports; from tumor registries; etc.)
Histology (i.e. from pathology reports; from tumor registries; etc.)
Morphology (i.e. from pathology reports; from tumor registries; etc.)
Grade (i.e. from pathology reports; from tumor registries; etc.)
Other cancer diagnoses (i.e., self-report; confirmed through medical record review; cancer registries, etc)
Other diseases (cardiovascular disease; auto-immune disease)
Death information (i.e., through social security index etc.)
Anthropometric variables (though self-report; measured by interviewers, etc.)
Pregnancies (i.e., queried number of children vs. number of pregnancies)
Analgesics (i.e., what was considered regular use)
Any other variables that you feel warrant further description?

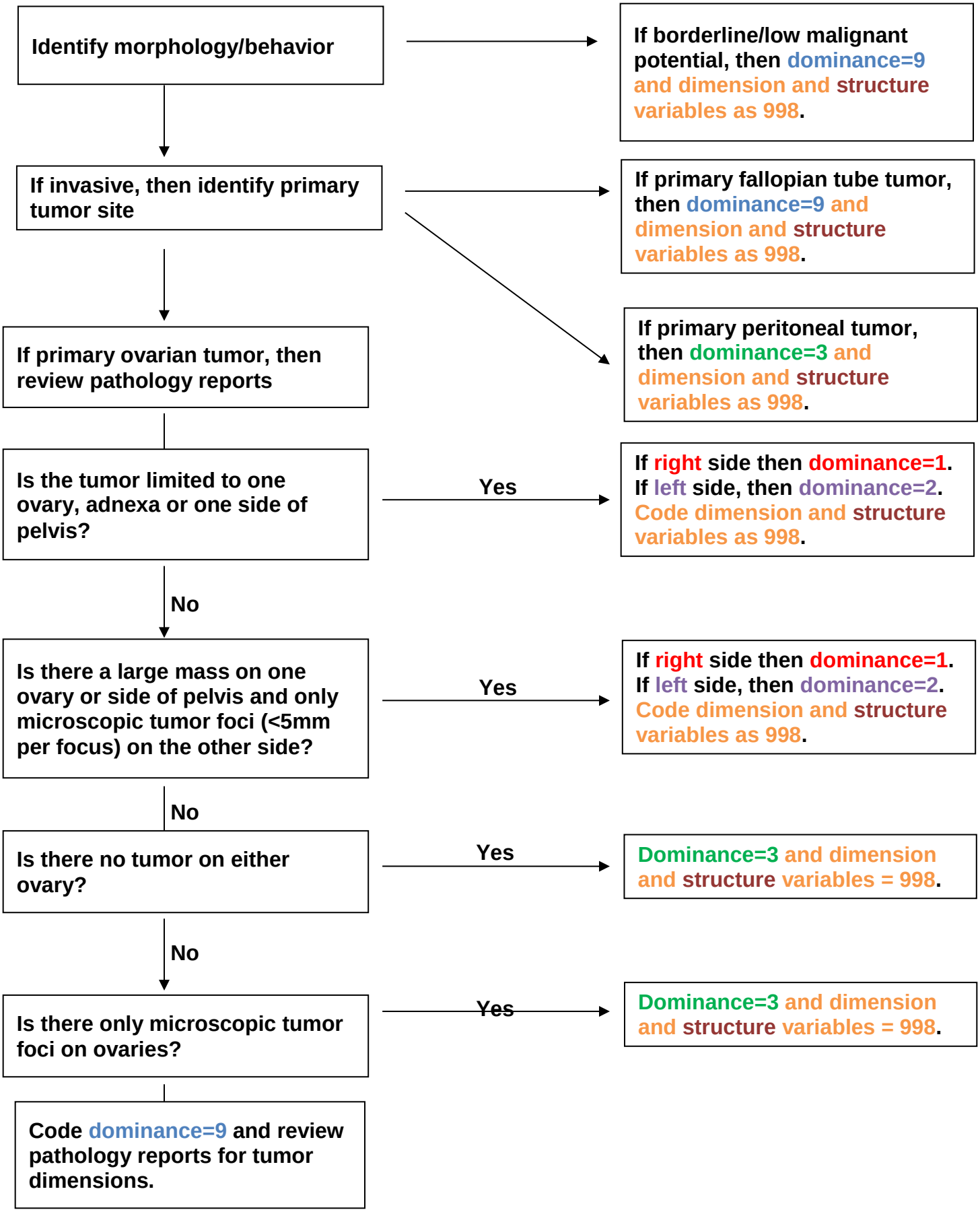
Part 2. Questions for future data requests

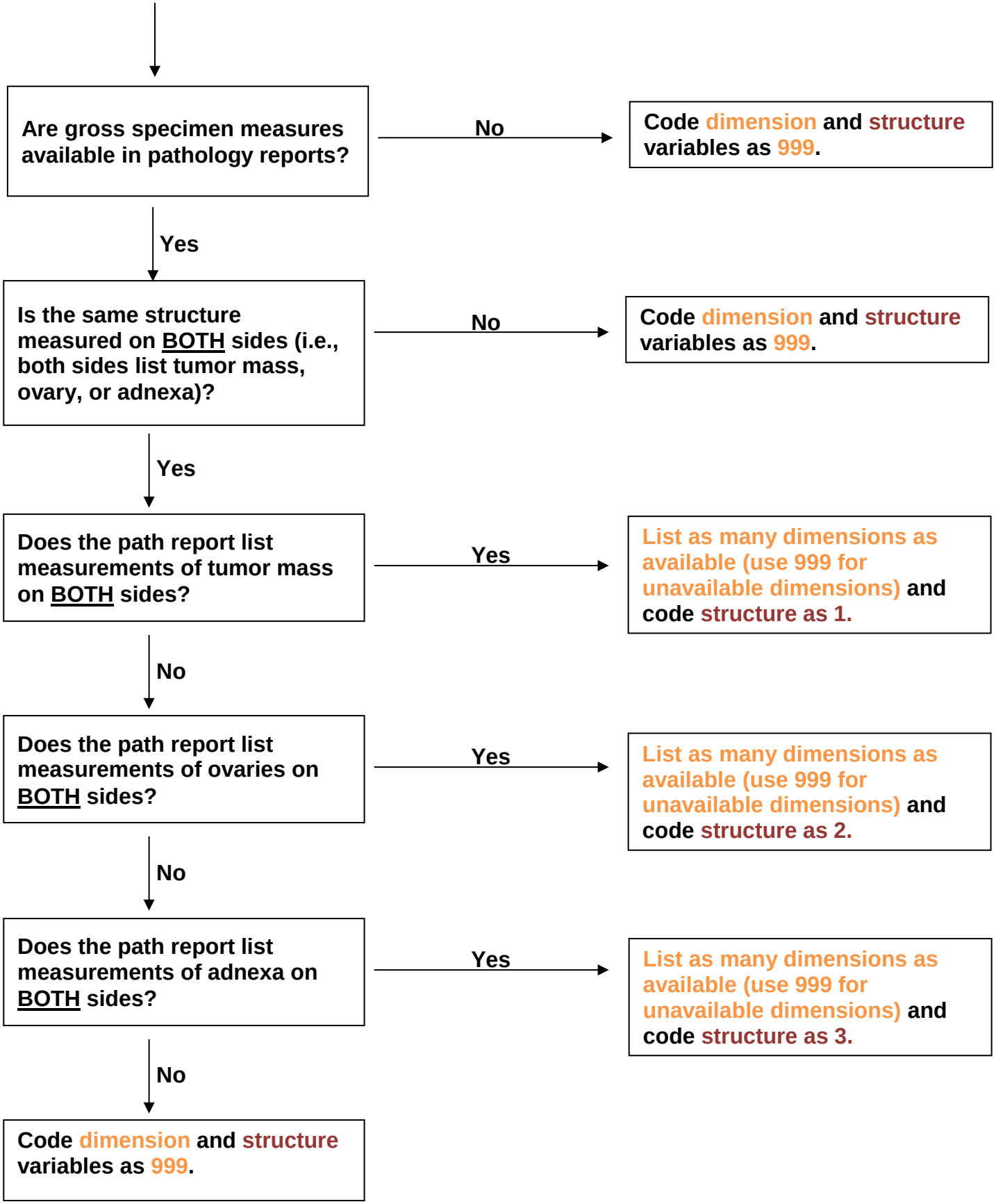
1. Do you have follow-up questionnaires after baseline? ___Yes ___No
    a. If yes, how many follow-up questionnaires have been sent and at what times?
b. Would you be willing to send follow-up data? ___Yes ___No
2. Do you have pathology reports? ___Yes ___No
    a. If yes, would you like for one of the OC3 staff to help you code ovarian tumor characteristics and/or outcomes? ___Yes ___No
3. Do you collect surgical reports? ___Yes ___No
4. Do you confirm additional cancers once a first cancer has been diagnosed and confirmed? ___Yes ___No

If yes, do you have ovarian tumors in women who have been diagnosed with other cancers either prior or subsequent to ovarian cancer diagnosis? ___Yes ___No

Appendix 2: Flow Chart for Coding Tumor Dominance

Variable Name	Description	Coding	Comment
STRUCTURE	Indicator of which ovarian structure was measured in pathology report	1=tumor; 2=ovary; 3=adnexa; 998=case was clearly dominant or non-dominant, no measurements required; 999=no measurements listed; missing if non-case	
DOMINANCE	Identifier of tumor dominance	1=dominant on right; 2=dominant on left; 3=non-dominant; 9=unknown; missing if non-case	
RDIM1	1st measurement of the tumor/ovary/adnexa on right side, if given	in cm, 998=case was clearly dominant or non-dominant, no measurements required; 999=no measurements listed; missing if non-case	if 3 dimensions not given, fill in as many as available in path report
RDIM2	2nd measurement of the tumor/ovary/adnexa on right side, if given	in cm, 998=case was clearly dominant or non-dominant, no measurements required; 999=no measurements listed; missing if non-case	
RDIM3	3rd measurement of the tumor/ovary/adnexa on right side, if given	in cm, 998=case was clearly dominant or non-dominant, no measurements required; 999=no measurements listed; missing if non-case	
LDIM1	1st measurement of the tumor/ovary/adnexa on left side, if given	in cm, 998=case was clearly dominant or non-dominant, no measurements required; 999=no measurements listed; missing if non-case	if 3 dimensions not given, fill in as many as available in path report
LDIM2	2nd measurement of the tumor/ovary/adnexa on left side, if given	in cm, 998=case was clearly dominant or non-dominant, no measurements required; 999=no measurements listed; missing if non-case	
LDIM3	3rd measurement of the tumor/ovary/adnexa on left side, if given	in cm, 998=case was clearly dominant or non-dominant, no measurements required; 999=no measurements listed; missing if non-case	


Identify morphology/behavior
If borderline/low malignant potential, then dominance=9 and dimension and structure variables as 998.
If invasive, then identify primary tumor site
If primary fallopian tube tumor, then dominance=9 and dimension and structure variables as 998.
If primary peritoneal tumor, then dominance=3 and dimension and structure variables as 998.
If primary ovarian tumor, then review pathology reports
Is the tumor limited to one ovary, adnexa or one side of pelvis?
Yes
If right side then dominance=1. If left side, then dominance=2. Code dimension and structure variables as 998.
No
Is there a large mass on one ovary or side of pelvis and only microscopic tumor foci (<5mm per focus) on the other side?
Yes
If right side then dominance=1. If left side, then dominance=2. Code dimension and structure variables as 998.
No
Is there no tumor on either ovary?
Yes
Dominance=3 and dimension and structure variables = 998.
No
Is there only microscopic tumor foci on ovaries?
Yes
Dominance=3 and dimension and structure variables = 998.
Code dominance=9 and review pathology reports for tumor dimensions.


Are gross specimen measures available in pathology reports?
No
Code dimension and structure variables as 999.
Yes
Is the same structure measured on BOTH sides (i.e., both sides list tumor mass, ovary, or adnexa)?
No
Code dimension and structure variables as 999.
Yes
Does the path report list measurements of tumor mass on BOTH sides?
Yes
List as many dimensions as available (use 999 for unavailable dimensions) and code structure as 1.
No
Does the path report list measurements of ovaries on BOTH sides?
Yes
List as many dimensions as available (use 999 for unavailable dimensions) and code structure as 2.
No
Does the path report list measurements of adnexa on BOTH sides?
Yes
List as many dimensions as available (use 999 for unavailable dimensions) and code structure as 3.
No
Code dimension and structure variables as 999.

Appendix 3: Proposal and publication guidelines for the OC3

Guidelines for proposing a new analysis using OC3 data

The OC3 is a rich source of questionnaire data relating to ovarian cancer risk and may also be useful in the future for molecular epidemiologic research. We anticipate that many of the OC3 participating studies will wish to propose new analyses. To ensure that these proposals are handled fairly and in a timely manner, we have established the following guidelines.

1. Before submitting a proposal to the cohort consortium (which is required by the NCI), email members of the steering committee (Shelley Tworoger: nhsst@channing.harvard.edu; Nico Wentzensen: wentzenn@mail.nih.gov; Liz Poole: nhlip@channing.harvard.edu) to make sure that your project has not already been proposed by another investigator. Ongoing projects and recently submitted proposals will also be posted on the OC3 website. It may also be useful to inquire how many cohorts have data on the exposure of interest. Also, tell the steering committee how the proposal will be funded.
2. Submit the proposal to the cohort consortium for approval (http://epi.grants.cancer.gov/documents/Consortia/Consortia_Project_Proposal_rev_4-1-11.pdf).
3. Once the cohort consortium has approved your proposal, please send the approved document to the steering committee. The steering committee will circulate approved proposals quarterly (i.e., once every 3 months) to the specified OC3 contact person for each cohort and ask that interested parties contact the proposal's PI to participate. Note that the OC3 operates on an opt-in basis. That is, cohorts have to officially give permission for their data to be used for each project. The Data Coordinating Center (DCC) requires an email from each cohort's contact person to the study PI or the DCC agreeing to participate in the analysis.
4. Once funding is secured, an analysis proposal must be written and provided to the DCC for review (to ensure data are available, etc.), to any cohorts requiring an analysis proposal submission, and for approval from the writing group.

Potential problems that may arise

1. What if two investigators propose the same project?

   If two investigators propose the same analysis within 6 months of each other, the steering committee will put the two investigators in touch to see if they can work together to come up with a proposal. However, if this does not work out, the first investigator to propose the analysis will be given precedence.

2. How long does each investigator have between proposing the analysis and getting started on the analysis?

   If an investigator proposes an analysis, but then does not attempt to secure funding (i.e., submit a grant proposal) within 1 year, then the steering committee will contact the investigator. If the investigator has no plans to submit a grant proposal (or otherwise secure funding), the steering committee will allow another OC3 member to propose that topic.

Ovarian Cancer Cohort Consortium (OC3) Publication Guidelines

Guidelines for authorship and publication have been developed to ensure that results published by the Ovarian Cancer Cohort Consortium (OC3) are timely and of the highest quality, that the research projects are effectively coordinated, and that the individual researchers and centers that contribute to the projects are given

fair and appropriate credit. These guidelines have been developed and approved by the OC3 Steering Committee. We anticipate that these guidelines will change over time to reflect the needs and experience of the OC3. At least during these early stages, the Steering Committee will review the guidelines at six-month intervals and propose modifications as needed.

General Principles

One of several objectives of the Cohort Consortium is to foster multidisciplinary collaboration among genomic researchers, epidemiologists, biostatisticians, and other scientists in large-scale prospective studies of ovarian cancer. We anticipate that the OC3 will generate many high impact publications. Recognizing the contribution of a large number of scientists to the collection and analysis of data in the ovarian cancer project, the OC3 has developed the following guidelines to coordinate and facilitate publications derived from this collaboration.

Authorship

Cohort	Center
BDCCP	NCI
Breakthrough Generations Study (BGS)	Institute of Cancer Research
California Teacher's Study (CTS)	City of Hope
Canadian Study of Diet, Lifestyle, and Health (CSDLH)	Albert Einstein College of Medicine
Cancer Prevention Study II (CPSII)	American Cancer Society
CLUE II	Johns Hopkins University
European Prospective Study into Cancer and Nutrition (EPIC) – 10 individual cohorts participating, not listed for brevity	Heidelberg/London coordinating centers
Iowa Women's Health Study	University of Minnesota
Melbourne Collaborative Cohort Study (MCC)	Cancer Council Victoria
Multi-Ethnic Cohort Study (MEC)	University of Hawaii
Netherlands Cohort Study	Maastricht University
NIH-AARP Diet and Health Study	NCI
Nurses' Health Study (NHS)	Harvard/Brigham and Women's Hospital
Nurses' Health Study II (NHSII)	Harvard/Brigham and Women's Hospital
NYU Women's Health Study (NYUWHS)	New York University
Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO)	NCI

Singapore Chinese Health Study	University of Pittsburgh
Sisters Study	NIEHS
Swedish Mammography Cohort	Karolinska Institute
Vitamins and Lifestyle Study (VITAL)	Fred Hutchinson Cancer Research Center
Women's Health Study (WHS)	Harvard/Brigham and Women's Hospital
Women's Lifestyle and Health Study (WLHS)	Karolinska Institute

The general approach to authorship will be inclusive rather than exclusive, although it should meet the criteria proposed by the International Committee of Medical Journal Editors (ICMJE) (Annals of Int Med 1988;258-304). The ICMJE specified that authorship credit should be limited to those who contribute substantially to all of the following:

a) conception and design, or acquisition of data, or analysis and interpretation of data:
b) drafting the article or revising it critically for important intellectual content:
c) final approval of the version to be published.

According to the ICMJE, none of the above contributions is sufficient by itself to justify authorship. However, in the case of the OC3, multiple centers have contributed to data collection by providing questionnaire and follow-up data on cohorts and/or by contributing to the statistical analyses. A center is defined as an institution that is participating in the OC3, while a cohort is defined as an individual study that resides at a center. Note that some centers will have multiple participating cohorts. Below is a list of the cohorts by center that have currently agreed to participate in the OC3:

- Currently 32 cohorts have agreed to participate from 18 centers.
- One center conducts data pooling and harmonization (Harvard/Brigham and Women's Hospital)

It is proposed that each cohort that participates in a particular analysis will have 2 co-authors (an author may represent more than one cohort), not counting the writing group, listed on that particular manuscript. Cohorts can request that the steering committee make an exception if more than two authors are required for certain cohorts. This approach is compatible with the ICMJE view that, "A test to help determine whether a center should be included in the author list is 'could the work have been completed without the center'? If not, then each center that provided data should be represented in the authorship."

The author list will begin with the members of the writing team (except the senior author). Listed after the writing team will be all other researchers (in alphabetical order) who meet the criteria for authorship. The last author will be the senior researcher responsible for that particular manuscript/project. Additionally, if the journal permits it, the writing group should be specified in a footnote or acknowledgement.

Each writing team will consist of five or six researchers who have worked on a specific exposure of interest or proposal and/or been most active in developing the OC3. When the senior or first author circulates a proposal to the participating cohorts for approval participation, he/she should also solicit participation on the writing team. Cohorts participating in the project can request that an individual from that cohort be on the writing team. The person who submitted the proposal will assemble the writing team from among those who expressed interest, spreading membership across the cohorts as much as possible. One researcher (generally determined by the senior researcher) on the writing team will take primary responsibility for data analysis and writing, and will be considered first author. The senior researcher will be the person who submitted the original proposal or is the mentor of the person who submitted the proposal; their role is to edit the paper and accept full responsibility for its content. Either the lead or senior investigator will serve as the corresponding author. The remaining order in which other authors on the writing team will be listed will be determined based on contribution to the writing team.

The duties of the writing team are as follows:

1. Advice on conducting data analysis
2. Assistance/insight in interpreting results
3. In-depth reading and feedback on draft manuscripts prior to circulation to the full group

These duties should be completed in a timely fashion to ensure that an analysis is completed within the timeframe that the OC3 has put forward (see below)

If major imbalances appear between groups, and if these imbalances do not fairly reflect the level of contribution, the authorship lists will be negotiated with the Steering Committee.

Final decisions about who should be considered full authors on collaborative papers will be made by members of the Steering Committee, not by the journal or PubMed/NML. Other mechanisms for acknowledgement are discussed below.

Acknowledgement of Other Contributions

All papers will acknowledge the source of funding as follows:

Dept. of Defense Ovarian Cancer Research Program (OC110197)

Any other sources of funding for participating studies will also be recognized after acknowledgement of the Consortium funding source. The Steering Committee will develop a list of acknowledgements for each cohort/center and post it on the website.

Manuscript preparation, review, and approval

A major challenge in large collaborative undertakings is to ensure timeliness and effective coordination in developing manuscripts. Given the funding period, timeliness is particularly important. Accordingly, we will establish a priority list of manuscripts, mechanisms for writing and incorporating feedback, and expected dates for submission that respects competing demands on collaborating members but maintains a pace that is appropriate for the priority of the undertaking.

The following checklist is proposed to encourage each writing team to designate responsibilities and to establish a timeline for manuscript development. A number of these steps can be addressed simultaneously rather than sequentially.

- Clearly delineate the role(s) of each member of the writing team
- Work with data coordinating center at Harvard/BWH to complete basic analyses; two possible options are: 1. The data coordinating center conducts all analyses under the direction of the first and senior authors or 2. The person conducting the analyses will be granted access to the data through the Harvard/BWH system. This will require modifying or setting up a new DUA to allow the analyst access to the data. The approach for each project should be discussed with Dr. Shelley Tworoger who oversees the data coordinating center.
- Develop outline of main analyses, table shells, and selection and format of figures
- Identify and conduct secondary analyses
- First draft of title, abstract, results, tables, & figures
- First draft of methods
- Complete literature review and draft of introduction and discussion.
- First draft of full manuscript
- Incorporate secondary analyses
- Review and revision by authors
- Review by Consortium

We propose the following guidelines for manuscript review by the overall Consortium.

- Each of the contributing centers or group of cohorts will identify a contact for coordinating their input to any given manuscript (i.e., the designated contributor). This contact is also responsible for obtaining any signatures or other paperwork necessary for final submission to the journal.
- Preliminary tables of results will be circulated along with the outline of the paper. When a draft manuscript is circulated, co-authors are expected to provide feedback within 2 weeks of receiving the draft. Once the first author has received feedback from the group, he/she has two weeks to incorporate the feedback and produce an updated version of the manuscript.
    - The submission of comments to the first author should be coordinated by the designated contributor from each cohort group, combining input before forwarding to the writing team.
    - Ideally comments will be ranked into two levels, (a) essential analytic and factual changes, and (b) possible grammatical and other editorial changes
    - Revisions of the manuscript will specifically address responses to the category (a) items noted above. (However, changes in category (b) above will also be accommodated as appropriate in the manuscript.)
- Revised manuscripts will be circulated to all co-authors with the understanding that a timely response is essential to the overall success of this research undertaking. Any further revisions should be returned to the first author within 2 weeks.
- The final manuscript will be submitted within six to eight weeks after any final comments are sent to the first author. This will allow for all necessary cohort-specific review processes (which will be listed on the website and in the table below) that must occur before submission. The first and senior authors are responsible for ensuring that these review processes are met.
- If comments are not sent within the appropriate time frame, these additional suggestions might be incorporated during the peer review process.

For this overall process to work smoothly, the writing team and all co-authors will be placing very high priority on the collaborative manuscripts. Clear communication and quick responses will be essential for success.

Appendix 4: Template data use agreements for participating cohorts and for those wishing to access the OC3 database with approved projects

Data Use Agreement Template for participating cohorts

This **Data Use Agreement** (the "Agreement") is by and between the _____("Holder") with its principal place of business in _____ and The Brigham and Women's Hospital, Inc. ("User") and is effective as of _____ (the "Effective Date").

WHEREAS, Congress enacted the Health Insurance Portability and Accountability Act of 1996 ("HIPAA"), which provides protection for confidential health information;

WHEREAS, the United States Department of Health and Human Services ("DHHS") promulgated, pursuant to HIPAA, a "Privacy Rule" governing confidential health information; "Privacy Rule" means the regulations promulgated by DHHS to implement the portions of HIPAA that concern the confidentiality of health information, as may be amended or otherwise changed from time to time. References to 45 CFR in this Agreement refer to sections within the Privacy Rule. At the date of this Agreement, these include 45 CFR §160 and §164, Subparts A and E;

WHEREAS, Holder maintains certain information that User wishes to use and /or disclose for research, public health, or health care operations purposes permitted under 45 C.F.R. §164.514(e) of the Privacy Rule; and

WHEREAS, prior to releasing any confidential health information to User, the Privacy Rule requires Holder to enter into an agreement under which User agrees to comply with the Privacy Rule.

NOW, THEREFORE, the parties, in consideration of the premises and the mutual promises and obligations set forth herein, the sufficiency of which is hereby acknowledged, and intending to be legally bound, agree as follows:

1. **Access to Data**. Holder shall provide User with access to certain data (the "Limited Data Set") in accordance with the terms and conditions of this Agreement. Under no circumstances shall Holder be required under this Agreement to provide the User with any information that does not qualify as part of a "limited data set" under 45 C.F.R. §164.514(e).

2. **Authorized Parties**. The following individuals (the "Authorized Parties") are authorized to use the Limited Data Set or any part of it on behalf of User and agree to abide by the terms of this Agreement:

Shelley Tworoger, Ph.D and her study staff

3. **Permitted Use**. User, and any Authorized Party on User's behalf, may use the Limited Data Set only for the following purposes (which shall be limited to research, public health activities, and health care operations as defined in the Privacy Rule):

For projects approved by the Ovarian Cancer Cohort Consortium Steering Committee and the NCI Cohort Consortium Secretariat for which the Holder has agreed to participate.

Use an attachment to list any additional purposes. The attachment must be signed by authorized representatives of User and Holder.

4. **Conditions of Use**. User and each Authorized Party agree as follows:

   a. Not to use or further disclose the Limited Data Set or any information contained therein other than as permitted by this Agreement or required by applicable law.

b. To use appropriate safeguards to prevent use or disclosure of the Limited Data Set or any information contained therein other than as provided for by this Agreement.

c. To report to Holder any use or disclosure of the Limited Data Set or any part of it not provided for by this Agreement of which User or any Authorized Party becomes aware.

d. To ensure that any agents, including subcontractors, and other members of the Ovarian Cohort Consortium to whom User or an Authorized Party provides the Limited Data Set or any part of it agree to the same restrictions and conditions that apply to the User and Authorized Parties under this Agreement.

e. Not to use the information contained in the Limited Data Set to identify the individuals whose information is contained in the Limited Data Set, nor to contact them under any circumstances.

f. Promptly following the end of the permitted use (as defined in Section 3 above), to return all copies of the Limited Data Set to Holder or destroy them and certify to the destruction; or, if User represents and Holder agrees that neither return nor destruction is feasible, to continue to extend the protections of this Agreement to the Limited Data Set. Notwithstanding the above, User may retain one copy of the Limited Data Set for verification purposes only consistent with the obligations and protections agreed to herein.

5. **Relief**. User and each Authorized Party agree that the breach or threatened breach of this Agreement may cause irreparable harm to Holder and/or individuals, that Holder may not have an adequate remedy at law, and that Holder may therefore be entitled to injunctive or other equitable relief to enforce this Agreement without obligation to post a bond. In the event Holder becomes aware of any use of the Limited Data Set or any part of it that is not authorized under this Agreement or required by applicable law, Holder may (i) terminate this Agreement upon notice; (ii) disqualify (in whole or in part) any Authorized Parties from receiving protected health information in the future; and/or (iii) report the inappropriate use or disclosure to the Secretary of the Department of Health and Human Services. Further sanctions may apply to the User and/or Authorized Parties under 45 C.F.R. parts 160 and 164.

6. **Obligations Following Termination**. Upon expiration or termination of this Agreement for any reason, User and all Authorized Parties shall no longer be entitled to receive or use information contained in the Limited Data Set.

7. **Expiration of Agreement**. Except as otherwise provided in Section 4.f. above, this Agreement shall expire thirty days following satisfaction of the requirements of Section 4.f. above.

8. **No Assignment**. This Agreement may not be assigned by User or any Authorized Party without the prior written consent of Holder.

WHEREFORE, the parties, through their authorized representatives, hereby accept and agree to the terms and conditions of this Agreement.

HOLDER	RECIPIENT
By:_____	By:_____
Name:_____	Name: Meghan Garland, J.D.
Title: _____	Title: Data Use Associate, Research Management
Date:	Date:

Data Use Agreement Template for those wishing to access the OC3 database for approved projects

This **Data Use Agreement** (the "Agreement") is by and between The Brigham and Women's Hospital, Inc. ("Holder") and LIST INSTITUTION ("User") and is effective as of _____ (the "Effective Date").

WHEREAS, Congress enacted the Health Insurance Portability and Accountability Act of 1996 ("HIPAA"), which provides protection for confidential health information;

WHEREAS, the United States Department of Health and Human Services ("DHHS") promulgated, pursuant to HIPAA, a "Privacy Rule" governing confidential health information; "Privacy Rule" means the regulations promulgated by DHHS to implement the portions of HIPAA that concern the confidentiality of health information, as may be amended or otherwise changed from time to time. References to 45 CFR in this Agreement refer to sections within the Privacy Rule. At the date of this Agreement, these include 45 CFR §160 and §164, Subparts A and E;

WHEREAS, Holder maintains certain information that User wishes to use and /or disclose for research, public health, or health care operations purposes permitted under 45 C.F.R. §164.514(e) of the Privacy Rule; and

WHEREAS, prior to releasing any confidential health information to User, the Privacy Rule requires Holder to enter into an agreement under which User agrees to comply with the Privacy Rule.

NOW, THEREFORE, the parties, in consideration of the premises and the mutual promises and obligations set forth herein, the sufficiency of which is hereby acknowledged, and intending to be legally bound, agree as follows:

1. **Access to Data**. Holder shall provide User with access to certain data (the "Limited Data Set") in accordance with the terms and conditions of this Agreement. All data will remain on the computer system managed by the Holder. Under no circumstances shall Holder be required under this Agreement to provide the User with any information that does not qualify as part of a "limited data set" under 45 C.F.R. §164.514(e). Obtaining individual computer access is dependent upon providing the Holder with proof of Human Subjects/IRB training and completing an application form to the Channing Division of Network Medicine for each individual who will access the data. Once a login is assigned, Authorized Parties will be assigned an individual directory (for each person completing the application) on which to write programs for analysis. This directory will allow the user to read the datasets and conduct appropriate statistical analyses. The User will not have direct access to the directories containing the datasets and as such will not be able to copy or alter the original data.

2. **Authorized Parties**. The following individuals (the "Authorized Parties") are authorized to use the Limited Data Set or any part of it on behalf of User and agree to abide by the terms of this Agreement:

LIST INDIVIDUALS

3. **Permitted Use**. User, and any Authorized Party on User's behalf, may use the Limited Data Set only for the following purposes (which shall be limited to research, public health activities, and health care operations as defined in the Privacy Rule) for a period not to exceed five (5) years from the Effective Date, unless otherwise agreed upon in writing by the parties:

To develop an ovarian cancer risk predication model that will be developed as part of the Ovarian Cancer Cohort Consortium and the Ovarian Cancer Association Consortium.

Use an attachment to list any additional purposes. The attachment must be signed by authorized representatives of User and Holder.

4. **Conditions of Use**. User and each Authorized Party agree as follows:

   a. Not to use or further disclose the Limited Data Set or any information contained therein other than as permitted by this Agreement or required by applicable law.

   b. To use appropriate safeguards to prevent use or disclosure of the Limited Data Set or any information contained therein other than as provided for by this Agreement.

   c. To report to Holder any use or disclosure of the Limited Data Set or any part of it not provided for by this Agreement of which User or any Authorized Party becomes aware.

   d. To ensure that any agents, including subcontractors, to whom User or an Authorized Party provides the Limited Data Set or any part of it agree to the same restrictions and conditions that apply to the User and Authorized Parties under this Agreement.

   e. Not to use the information contained in the Limited Data Set to identify the individuals whose information is contained in the Limited Data Set, nor to contact them under any circumstances.

   f. Promptly following the end of the permitted use (as defined in Section 3 above), access to the data and computer system will be revoked; or, if User represents and Holder agrees that continued access is needed to extend the protections of this Agreement to the Limited Data Set that will be discussed.

   g. User agrees to acknowledge the contribution of the National Cancer Institute, Division of Cancer Epidemiology and Genetics (DCEG) in all written or oral public disclosures concerning User's research using the Limited Data Set, as is appropriate. User agrees to provide Holder with copies of public materials based on use of the data.

5. **Relief**. User and each Authorized Party agree that the breach or threatened breach of this Agreement may cause irreparable harm to Holder and/or individuals, that Holder may not have an adequate remedy at law, and that Holder may therefore be entitled to injunctive or other equitable relief to enforce this Agreement without obligation to post a bond. In the event Holder becomes aware of any use of the Limited Data Set or any part of it that is not authorized under this Agreement or required by applicable law, Holder may (i) terminate this Agreement upon notice; (ii) disqualify (in whole or in part) any Authorized Parties from receiving protected health information in the future; and/or (iii) report the inappropriate use or disclosure to the Secretary of the Department of Health and Human Services. Further sanctions may apply to the User and/or Authorized Parties under 45 C.F.R. parts 160 and 164.

6. **Obligations Following Termination**. Upon expiration or termination of this Agreement for any reason, User and all Authorized Parties shall no longer be entitled to receive or use information contained in the Limited Data Set.

7. **Expiration of Agreement**. Except as otherwise provided in Section 4.f. above, this Agreement shall expire thirty days following satisfaction of the requirements of Section 4.f. above.

8. **No Assignment**. This Agreement may not be assigned by User or any Authorized Party without the prior written consent of Holder.

WHEREFORE, the parties, through their authorized representatives, hereby accept and agree to the terms and conditions of this Agreement.

HOLDER RECIPIENT

By:_____
Name: Meghan Garland, J.D.
Title: Senior Agreement Associate
Date:

By:_____
Name:
Title:
Date:

Table 1: Participating cohorts in the OC3 and estimated numbers of women and ovarian cancer cases as well as the status of the data use agreement (DUA) and receipt of data to the data coordinating center (DCC)

Cohort	Cohort Size*	Ovarian Cancer Cases	DUA signed^	Data received at DCC
Breast Cancer Detection Demonstration Project Follow-up Study (BCDDP)	38,697	205	Yes	Yes
Breakthrough Generations Study	111,000	330	No	No
California Teacher's Study (CTS)	44,200	236	Yes	Yes
Canadian Study of Diet, Lifestyle, and Health	39,618	142	No	No
Cancer Prevention Study II (CPS2)	94,595	660	Yes	Yes
CLUE II: Campaign Against Cancer and Heart Disease* (CLUE)	12,406	88	Yes	Yes
EPIC	274,442	1,035	No	No
Iowa Women's Health Study (IWHS)	30,716	361	Yes	Yes
Melbourne Collaborative Cohort Study	24,000	136	Yes	No
Multi-Ethnic Cohort (MEC)	17,664	107	Yes	Yes
Netherlands Case-Cohort Study	2,896	484	Yes	Yes
NIH-AARP Diet and Health Study (AARP)	225,388	1000	Yes	Yes
Nurses' Health Study (NHS FU1 & FU2)			NA	Yes
Follow up period from 1980-1996	103,298	547		
Follow up period from 1996-2010	100,393	590		
Nurses' Health Study II (NHS2)	116,430	281	NA	Yes
NYU Women's Health Study (NYUWHS)	12,472	138	Yes	Yes
Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO)	65,090	444	Yes	Yes
Singapore Chinese Health Study	34,028	121	Yes	Yes
Sisters Study	50,884	300	Yes	Yes
Swedish Mammography cohort	61,057	301	Yes	No
Vitamins and Lifestyle Study (VITAL)	33,252	163	Yes	Yes
Women's Health Study (WHS)	39,876	225	NA	Yes
Women's Lifestyle and Health Study	49,258	375	Yes	Yes
Total	**1,581,660**	**8,269**		

*Numbers based on data received when available or estimations from study specific principal investigator
^For those studies with no DUA, negotiations are on-going

Figure 1: Schematic for developing and validating a risk prediction model of ovarian cancer in the OCAC and OC3.
P(BL) = the baseline risk of developing ovarian cancer from SEER
P(BSO|A) = the probability of having bilateral oophorectomy at age, A from NHANES
P(RR|BSO) = the risk of developing ovarian cancer after bilateral oophorectomy, from a literature review

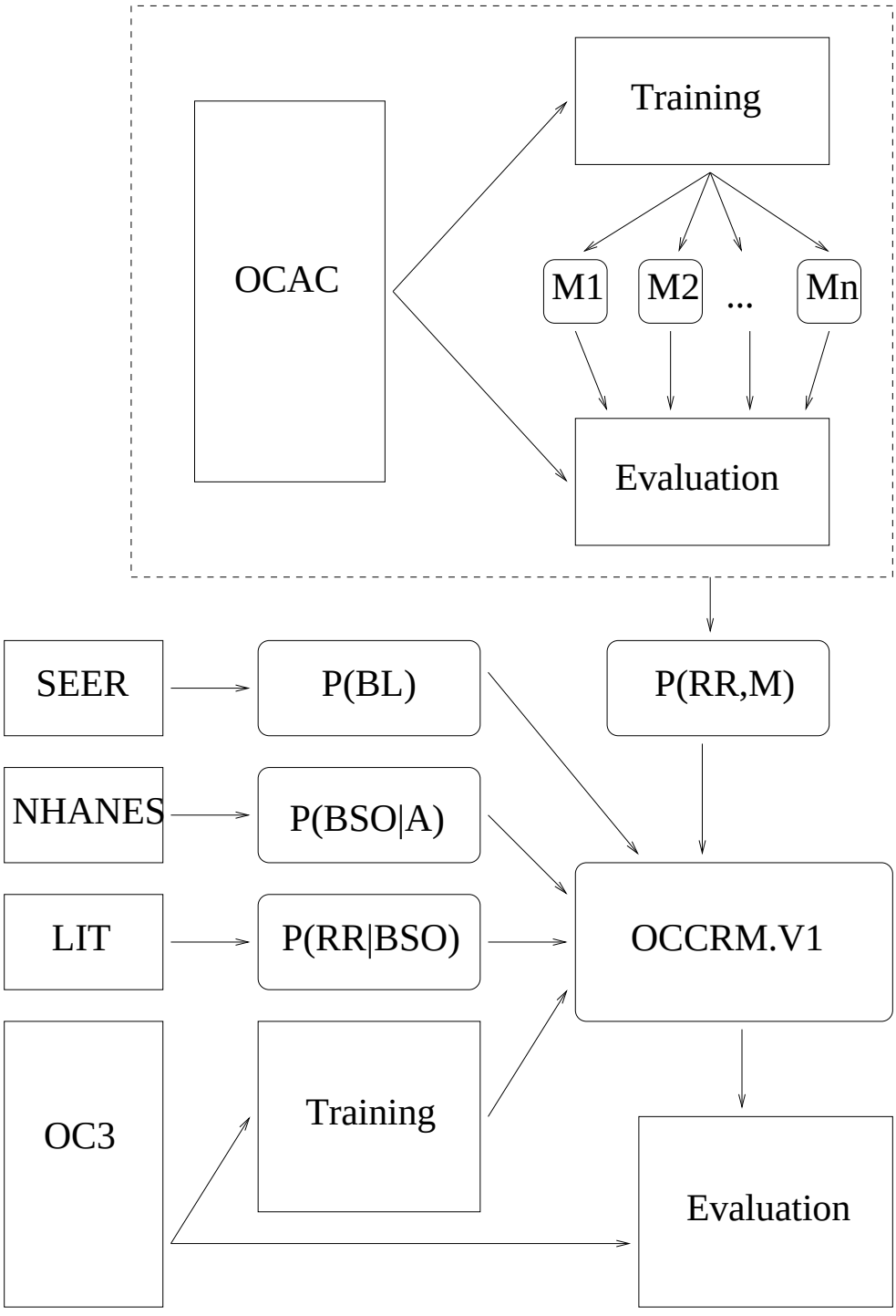


Table 2: Preliminary results for the association of oral contraceptive use and parity with total ovarian cancer risk among 13 of the cohorts participating in the OC3

	Oral contraceptive use		Parity	
Cohort	**RR, ever vs. never**	**RR, duration per year**	**RR, ever vs. never**	**RR, per child**
AARP	0.72	0.95	0.69	0.91
BCDDP	0.81	0.96	1.04	0.95
CPS2	0.78	0.96	0.79	0.93
CLUE	0.67	1.00	0.96	0.83
PLCO	0.93	0.98	0.83	0.92
NHS FU1	0.81	0.94	0.49	0.86
NHS FU2	0.83	0.98	0.73	0.93
NHS2	1.04	1.00	0.57	0.76
MEC	0.87	0.96	0.83	0.95
WHS	1.02	1.01	0.56	0.88
VITAL	1.08	1.00	1.02	0.95
NYUWHS	0.75	0.97	NA*	0.89
CTS	0.81	0.96	0.77	0.90
IWHS	0.97	0.97	NA*	0.92
Meta-analysis	**0.88 (0.80, 0.96)**	**0.98 (0.97, 0.99)**	**0.71 (0.62, 0.80)**	**0.90 (0.88, 0.93)**
Q-statistic (p)	23.9 (0.03)	47.8 (p<0.001)	17.5 (0.09)	19.6 (0.11)

*All women in the study were parous thus a risk estimate could not be determined for these studies

Table 3: OC3 Steering Committee members

Name	Institution
Shelley Tworoger (Co-Chair)	Brigham and Women's Hospital
Nicolas Wentzensen (Co-Chair)	National Cancer Institute
Alan Arslan	New York University
Laura Baglietto	University of Melbourne
Lesley Butler	Colorado State University
Annakatrin Lukanova	German Cancer Research Centre (Heidelberg)
Stefanie Nelson (Advisor)	National Cancer Institute
Alpa Patel	American Cancer Society
Elizabeth Poole	Brigham and Women's Hospital
Sheri Schully (Advisor)	National Cancer Institute
Kala Visvanathan	Johns Hopkins University
Emily White	Fred Hutchinson Cancer Research Center

Table 4: Additional proposed projects using the OC3 infrastructure and the status

Project Name	Proposed by	Date	Status
NSAIDs and ovarian cancer risk	Wentzensen, Trabert	Feb. 2012	Requested NCI internal funding
Endometriosis and ovarian cancer risk	Wentzensen, Trabert	Feb. 2012	Requested NCI internal funding
CRP/inflammatory risk factors and ovarian cancer risk	Poole, Tworoger	Nov. 2012	NCI R03 second submission Nov 2013
Androgens and ovarian cancer risk	Lukanova	Oct. 2012	EU application submitted
IGFs and ovarian cancer risk	Lukanova	Oct. 2012	EU application submitted
Diabetes and ovarian cancer risk	Patel, Gapster	Nov. 2012	Obtaining funding
OncoArray (GWAS)	Wentzensen, Tworoger	June 2013	NCI U19 funding (PI: Sellers)